Joint Sparsity in SAR Tomography for Urban Mapping

Xiao Xiang Zhu, Nan Ge, Muhammad Shahzad

***Abstract*—With meter-resolution images delivered by modern SAR satellites like TerraSAR-X and TanDEM-X, it is now possible to map urban areas from space in very high level of detail using advanced interferometric techniques such as PSI and SAR Tomography (TomoSAR), whereas these multi-pass techniques are based on a great number of images. We aim at precise TomoSAR reconstruction while significantly reducing the required number of images by incorporating building a priori knowledge to the estimation. In the paper, we propose a novel workflow that marries the freely available 2-D building footprint GIS data and the joint sparsity concept for TomoSAR inversion. Experiments on bistatic TanDEM-X data stacks demonstrate the great potential of the proposed approach, e.g., highly accurate tomographic reconstruction is achieved using six interferograms only.**

***Index Terms*—Synthetic aperture radar, SAR tomography, compressive sensing, joint sparsity, GIS, TanDEM-X.**

I. Introduction

MODERN spaceborne synthetic aperture radar (SAR) sensors, such as TerraSAR-X, TanDEM-X and COSMO-SkyMed, deliver SAR data with a very high spatial resolution (VHR) of up to 1 m. With these meter resolution data, advanced multi-pass interferometric techniques such as persistent scatterer interferometry (PSI) and tomographic SAR inversion (TomoSAR) allow retrieving not only the 3-D geometrical shape but also the undergoing temporal motion in the scale of millimeter of individual building [1][2][3][4][5][6][7]. In particular, sparse reconstruction based methods [8][9], like SL1MMER [10], give robust TomoSAR inversion with very high elevation resolution, and can offer so far ultimate 3-D, 4-D and 5-D SAR imaging [11][12].

The downside of advanced repeat-pass InSAR techniques [13][14][15][16][17][18][19][20], like PSI and TomoSAR, are their high demand on the data, i.e., typically a stack of 20–100 images over the illuminated area are required. For instance, it is demonstrated in [10] that even using most efficient algorithms, like non-linear least squares (NLS) and SL1MMER, a minimum number of 11 acquisitions are required to achieve a reasonable reconstruction in the interesting parameter range of spaceborne SAR. "Reasonable" in this context means that given an average signal-to-noise ratio (SNR) of 6 dB, the detection rate of double scatterers with an elevation distance of one Rayleigh resolution unit reaches at least 90%. However, if we can extract certain detailed features or patterns of high-rise buildings in SAR images, the required number of images can be significantly reduced by incorporating such features as prior for a joint estimation.

For this purpose, we propose a novel workflow marrying the globally available (2-D building footprint) GIS data and the joint sparsity concept for TomoSAR inversion, both of which have not yet been addressed in the community so far. Within this workflow, our main contributions are as follows:

- A robust procedure is proposed to use online freely assessable 2-D building footprints for extracting detailed high-rise building features including building masks, orientation, and iso-height lines (defined in [21]) in SAR image stacks (see Section III);
- The M-SL1MMER algorithm is proposed to promote joint sparsity for tomographic inversion of the identified iso-height pixel groups (see Section IV);
- The performance of M-SL1MMER is systematically evaluated using simulated data in terms of elevation estimation accuracy, detection rate and false alarm rate of the overlaid scatterers' separation, and its super-resolution capability. Compared to the single-snapshot sparsity model, as used in SL1MMER, the superior performance of the proposed joint sparsity approach is evident for all above mentioned quantitative metrics (see Section V);
- The first tomographic reconstruction using bistatic TanDEM-X data stacks is presented. The superior performance of M-SL1MMER is demonstrated in practice, e.g., highly accurate tomographic reconstruction is achieved using six interferograms only (see Section Section VI).

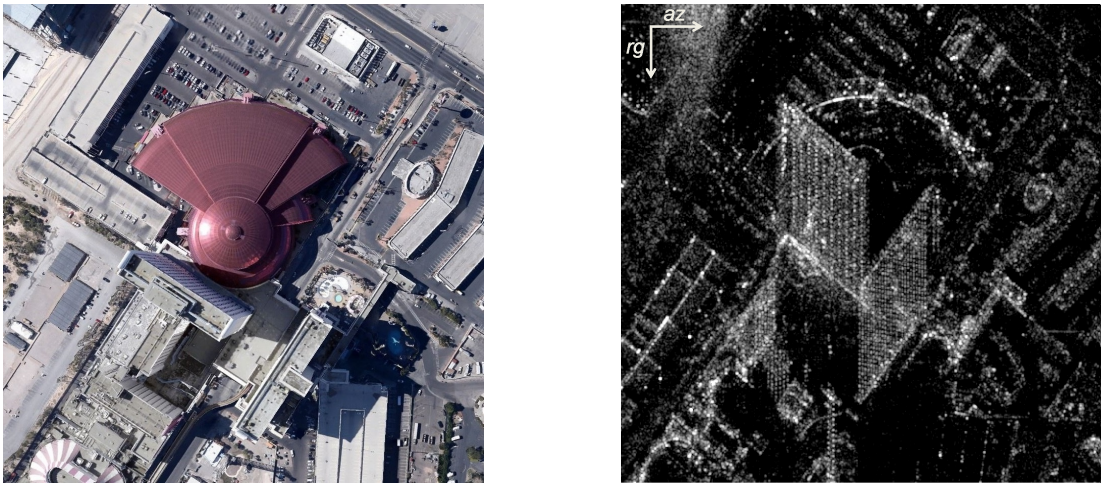
II. Data Set

We work with 21 bistatic interferograms acquired by the German SAR satellites TerraSAR-X and TanDEM-X, with cross-track baselines ranging between approximately ± 200 [m]. The single-pass characteristic renders atmospherical effects very small and deformation negligible. For this reason these datasets are ideal to test our proposed methodology. An optical image of the test area is shown in Figure 1(a) while the corresponding SAR mean intensity image is shown in Figure 1(b).

X. X. Zhu is with Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, 82234 Wessling, Germany and Signal Processing in Earth Observation (SiPEO), Technische Universität München, 80333 Munich, Germany (E-mail: xiao.zhu@dlr.de).

N. Ge is with Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, 82234 Wessling, Germany (E-mail: nan.ge@dlr.de).

M. Shahzad is with Signal Processing in Earth Observation (SiPEO), Technische Universität München, 80333 Munich, Germany (E-mail: muhammad.shahzad@tum.de).



(a) (b)

Fig. 1. Test Area: (a) Optical image of the test area ©Google; (b) Corresponding SAR intensity map (rg and az refer to the range and azimuth coordinates respectively).

III. Prior Knowledge Retrieval

In order to retrieve prior information pertaining to building regions, the 2-D building footprints are downloaded from OpenStreetMap (OSM). Based on the concept of crowd sourcing that involve crowd or community to effectively and efficiently fulfill a task at hand, OSM with around 2 million registered users (as of today and also rapidly growing) is considered to be the most successful Volunteered Geographic Information (VGI) project [22][23]. The OSM database contains multitude of building footprints represented as polygons with ordered list of nodes/vertices (i.e., pairs of UTM or latitude/longitude coordinates according to WGS 84 coordinate system) and is updated every day. The data are free to download and comes under the open license Open Data Commons Database License (ODbL). Since it is a VGI project, the data quality may vary from region to region. To this end, the first investigations regarding OSM data quality were carried out for roads [24] followed by assessment of other attributes present in the database e.g., lines [25], polygonal objects [26] etc. Recently, the building footprints have also been evaluated for their completeness [27] and correctness [23]. The analysis of OSM data with surveying datasets reveals fairly precise positioning accuracies varying within 4 meters [23][24]. The completeness percentage is already very high for many cities in Europe and US and is consistently increasing with time. Available 2-D footprints of the buildings in the city of Las Vegas are shown in Figure 2 to give the reader an insight of the existing database. The high availability of such type of data triggers us to change our perspective of thinking, namely, instead of using Earth observation (EO) satellite data to build-up sources of geo-information for open users, we can explore the knowledge provided by social media to support information retrieval from EO data. In this regard, one mission of this work is to demonstrate this concept in tomographic SAR reconstruction.

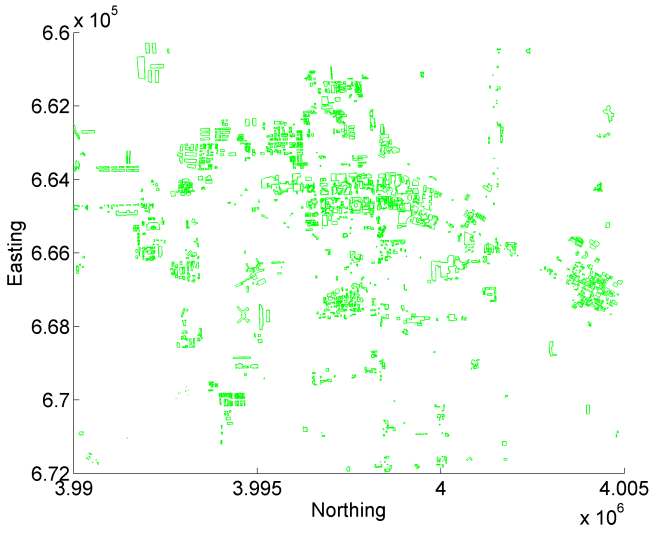


Fig. 2. GIS data (2-D building footprints) of Las Vegas from OSM.

A. Automatic extraction of building mask in SAR image

The key idea is to make use of the aforementioned online freely assessable 2-D building footprints to extract detailed high rise building features including building masks, orientations, and the iso-height lines in SAR image data stacks. The extracted information can be further incorporated as a prior knowledge into the estimation for a more accurate tomographic SAR inversion. For this purpose, in this section we propose a sophisticated approach that is tolerable to moderate errors in the input GIS data for automatic extraction of aforementioned high rise building features in the SAR image data stacks:

- First the available building footprints from OpenStreetMap in world (latitude/longitude) coordinates are transformed/geo-coded into SAR (azimuth/range) imaging coordinate system. Figure 3(a) shows the resulting projected reference polygons overlaid onto the buildings

of interest in the corresponding SAR image shown in Figure 1;

- Secondly, due to the side looking geometry, SAR illuminates only one side of the building. Therefore, the complete building footprint of individual buildings is further segmented into two parts by means of a simple 2-D visibility test: 1) the part illuminated by the sensor which will be further used for iso-height pixel extraction (as depicted by red polylines in Figure 3(a)); 2) the part in the shadow area not visible to the sensor which will not used in further processing (as depicted by green polylines in Figure 3(a));
- Finally, errors in the identified red polylines, caused by inaccuracies of the input GIS data in both orientation and translation, are compensated and the mask of individual buildings is further generated by iteratively shifting the corrected polylines towards the sensor.

In this regard, the approach depicted in Algorithm 1 is adopted. After transforming the available building footprints from world coordinates to SAR imaging coordinate system, we identify the side of the building footprint facing the SAR sensor as follows. If we assume that $v_{i=1,\ldots,n}$ denote the indices of ordered 2-D footprint vertices of one particular building. Then any vertex $v_k(k \in n)$ belongs to the side facing the sensor if and only if its projection onto the line at zeroth range axis (i.e., line defined as $rg = 0$ with zero azimuth slope) does not self-intersect the reference polygon. The range of total number of vertices belonging to the side visible to the sensor in any footprint is m where $1 < m \leq n$. The inequality that $m > 1$ depicts that, if not occluded, at least one side or two vertices of the building are always visible to the side looking SAR sensor.

Once the vertices facing to the sensor are identified, the step 3 in Algorithm 1 compensates for any positioning inaccuracy in the OSM footprint of the building in the area of interest. Possible error in OSM footprint is compensated by adopting the following sequence of steps:

1) Shift/translate the identified polyline in 2-D sliding window fashion within the intervals: range shift $[-10\ 10]$ and azimuth shift $[-5\ 5]$;
2) Within each shift, rotate the polyline between interval $[-7.5\ 7.5]$ degrees and compute median of intensities along the rotated polyline (similar to steps 6-7 in Algorithm 1);
3) The polyline is rotated and shifted with the rotation angle and the azimuth-range shifts which give the maximum of computed median intensities (from the previous step);
4) Finally, the rotation causes the change in the azimuthal length of the polyline which needs to be adjusted (see Figure 4(a)). This is accomplished by first slightly extending the polyline and later adjusting the lengths of the outer (first and the last) edges of the polyline by analysing their (interpolated) intensities (same as step 2 except that only the first and last edge of the polyline is used). Figure 4(b) graphically illustrates the adjustment procedure.

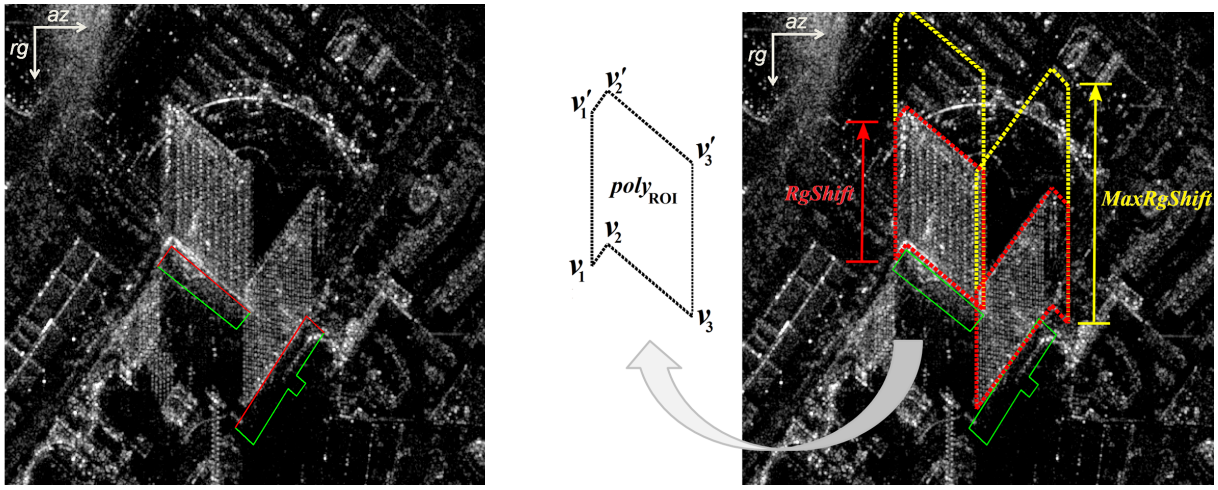
Algorithm 1 Procedure to automatically generate mask (or ROI) of an individual building

Require: 2-D polygonal footprint vertices $\mathbf{v}_{i=1,\ldots,n}$ in SAR coordinates of one particular building & SAR image of the scene.

1: Initialize: $MaxRgShift := 595$ and $d := 1$
2: Identify the polyline comprising of m out of n vertices belonging to the building side facing the sensor
3: Apply orientation correction and range-azimuth shift compensation to the identified polyline in order to cope for any positioning inaccuracy in the OSM footprint of the building
4: **while** (1) **do**
5: Shift/translate (in range direction) the polyline after orientation correction and azimuth-range compensation towards the sensor by distance d
6: Compute the intensity values along the shifted polyline. This is accomplished by selecting equally spaced points along the shifted polyline, and then using nearest neighbor interpolation to find the intensity value for each point
7: Take the median of computed intensity values along the shifted polyline and store the result in a column matrix $C(d, 1)$
8: **if** $d == MaxRgShift$ **then**
9: $break$
10: **else**
11: $d := d + 1$
12: **end if**
13: **end while**
14: Take approximate derivative of C (i.e., calculate differences between adjacent elements of C), and store the result in matrix D
15: Compute $\arg\max_{RgShift}(D_{RgShift})$ where $RgShift$ $(= 1, \ldots, d-1)$ denotes the maximum change point in D
16: Use $RgShift$ and m vertices of the identified (compensated) polyline facing the sensor to extract $poly_{ROI}$ of the particular building

In our experiment, the shifting intervals used in compensation are based on already mentioned 4m inaccuracy of OSM data i.e., considering 4m inaccuracy, approx. range and azimuth resolution of 0.588m and 1.1m requires range and azimuth shifts of atmost $(4/0.588 \approx)$ 6.8 pixels and $(4/1.1 \approx)$ 3.6 pixels.

After compensating any orientation and/or shifting inaccuracies, steps 4 to 13 in Algorithm 1 iteratively shifts the compensated polyline towards the sensor (in range direction only). Since the tallest building in the city of Las Vegas, the Stratosphere Tower, is around 350 m, $MaxRgShift$ in Algorithm 1 is set to 595 i.e., maximum building size along elevation appearing in the SAR image of Las Vegas city, computed as $350/0.588(\approx 595$ pixels) where 0.588m is the approx. range resolution. Thus the polyline is shifted till $MaxRgShift$ and the column vector C stores the median



(a) (b)

Fig. 3. Building mask extraction: (a) Reference polygons (shown in red and green polylines) of two buildings in the area of interest overlaid onto the SAR intensity map after geocoding. Side of the buildings facing the sensor are shown in red while the other side not visible to the sensor in green; (b) After rotation and range-azimuth shift compensation, the red polylines in (a) are shifted towards the sensor. The yellow dotted lines indicate the maximum range shift of $MaxRgShift$ (= 595m) where as the red dotted lines indicate the $RgShift$ obtained by the proposed procedure in Algorithm 1.

of computed intensity values along each range shift. Steps 14 and 15 in Algorithm 1 then computes the maximum change point $RgShift$ in the approximate derivative of C. $RgShift$ is then used in step 16 to determine $poly_{ROI}$ which describes the polygon surrounding the overlaid pixels of the same building in the SAR image. To elaborate how $poly_{ROI}$ is computed, consider a building having three adjacently connected vertices $v_1 - v_2 - v_3$ of the polyline facing the sensor where $-$ denotes the adjacency (i.e., v_2 is adjacently connected to v_1 and v_3, and so on). Assuming that the polyline has been compensated for rotation and range-azimuth shifts, the $poly_{ROI}$ is then simply formed as $v_1 - v_2 - v_3 - v_3' - v_2' - v_1' - v_1$ where $\mathbf{v}_j'(az, rg) = \mathbf{v}_j(az, rg - RgShift)$ with $j = 1, 2, 3$ (see Figure 3(b)). Finally, $poly_{ROI}$ is used to generate the building mask of an individual building.

B. Pixel grouping

Based on the extracted masks of individual buildings, pixels sharing similar heights are then grouped together. This procedure is done in three steps:

1) Iso-height lines will be reproduced by translating the adjusted polyline towards both ends of the building mask with sub-pixel increments;
2) The distance between each pixel and its adjacent iso-height lines will be calculated;
3) Each pixel will be assigned to the closest iso-height line.

The distance between a pixel and an iso-height line is defined as the minimum absolute amount of translation (in pixels) towards or away from sensor. Figure 5 illustrates one exemplary iso-height line in the cropped intensity image, as well as the final results of pixel grouping with each group of pixels plotted with a random color. Note that the color-coding already gives a rough idea about monotonic height change of the investigated façades.

IV. Joint Sparsity in TomoSAR

In this section, we first revisit a data model commonly used in TomoSAR, as well as the SL1MMER algorithm. Following this, we extend the SL1MMER algorithm for the multiple-snapshot case. The extended version exploits joint sparsity and is named as M-SL1MMER.

A. TomoSAR system model

For a single SAR image, information along the third dimension, the so-called elevation axis s, which is perpendicular to the azimuth-range (x-r) plane, is integrated (see Figure 6). I.e., echoes from, e.g., tree crown, building roof, or double-bounce effects on a balcony sharing the same distance to the sensor, are mapped onto one single pixel. To reconstruct reflectivity along s and to further separate those different contributions, TomoSAR utilizes scenes acquired from slightly different viewing angles to synthesize an elevation aperture Δb (cf. aperture along x created by steering the radar beam) for full 3-D SAR imaging [28]. A well-established model, which can be found, e.g., in [29], approximates each pixel value g_n as follows:

$$g_n = \int_{\Delta s} \gamma(s) \exp(-j2\pi\xi_n s)\, \mathrm{d}s, \tag{1}$$

which is essentially the Fourier transform of the reflectivity function $\gamma(s)$ sampled at the spatial (elevation) frequency $\xi_n = -2b_n/(\lambda r)$, Δs is the elevation extent and λ is wavelength. Note that a possible motion term has been neglected here without loss of generality. For differential SAR tomography that takes into account the motion component, the readers are recommended to consult [17][30][31].

Discretizing the continuous function in Eq. (1) along elevation s into L elevation positions $s_l \, \forall \, l \in \{1, \ldots, L\}$, replacing integral by sum and taking into account measurement noise yield the following discrete system model:

$$\mathbf{g} = \mathbf{R}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{2}$$

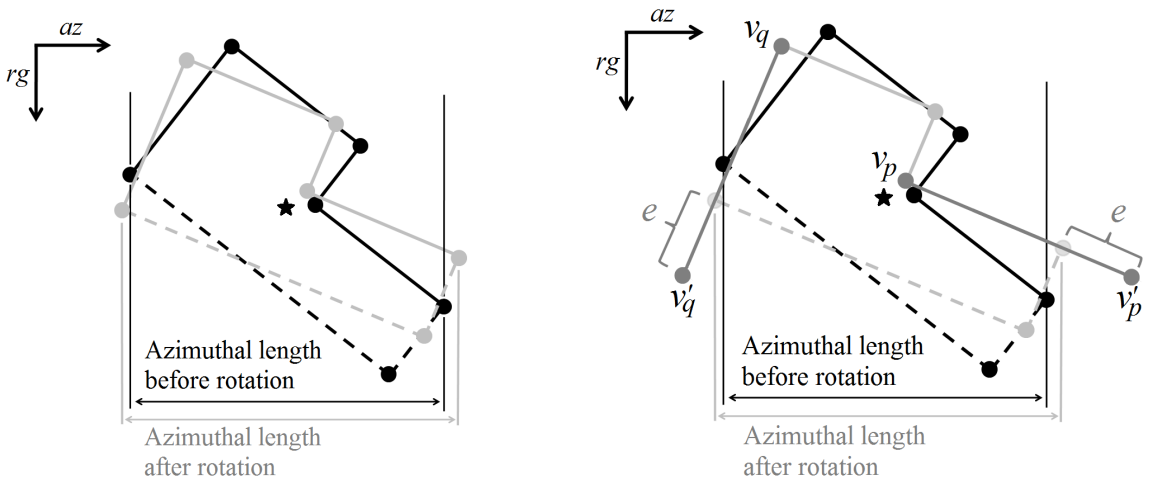


Fig. 4. Graphical illustration of adjusting polyline length: (a) Black and gray polygons indicate polygons before and after rotation (around centroid depicted as black star) respectively with circles representing the corresponding vertices. The dotted polylines represent building side not visible to the sensor. It is shown that after rotation the azimuthal length is changed; (b) The length of the outer (first and last edge depicted in dark gray) edges of the polyline facing the sensor is slightly extended by distance e (= 5m in this work). Intensities (interpolated) over these extended edges are analysed and first and last extended points (i.e., v'_p and v'_q) are respectively replaced by the points on the edges $v_p v'_p$ and $v_q v'_q$ where the approximate derivative (or change in intensities) is maximum.

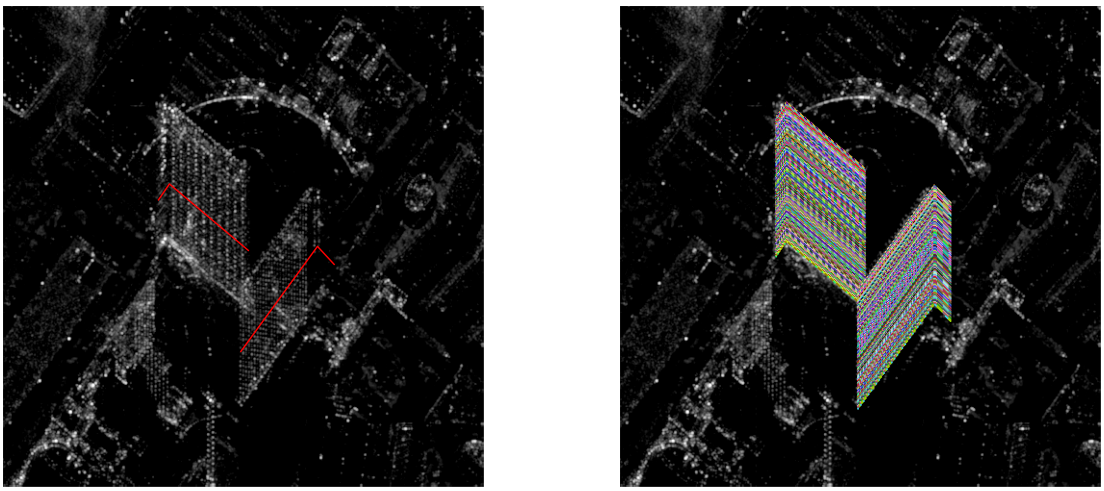


Fig. 5. Illustration of pixel grouping with (a) exemplary iso-height lines, and (b) grouped iso-height pixels color-coded with group indices.

where $\mathbf{g} \in \mathbb{C}^{N\times 1}$ is the measurement vector with $g_n \,\forall\, n \in \{1,\ldots,N\}$, $\mathbf{R} \in \mathbb{C}^{N\times 1}$ is an irregularly sampled Fourier transform matrix with $R_{nl} = \exp(-j2\pi\xi_n s_l)$, $\boldsymbol{\gamma} \in \mathbb{C}^{L\times 1}$ is the discretized reflectivity vector $\gamma_l = \gamma(s_l) \,\forall\, l \in \{1,\ldots,L\}$, and $\boldsymbol{\varepsilon} \in \mathbb{C}^{N\times 1}$ is additive noise which can be modeled as a zero-mean circular Gaussian random process. Typically we have $N \ll L$, which renders Eq. (2) underdetermined.

Similar to the resolution in azimuth, the Rayleigh resolution ρ_s is inversely proportional to the aperture size [2]

$$\rho_s = \frac{\lambda r}{2\Delta b}. \tag{3}$$

For high resolution spotlight data of TerraSAR-X/TanDEM-X, ρ_s is much worse than azimuth and range resolution (approx. 1.10 and 0.588 m, respectively) due to tight orbit control and amounts to about 24.9 m for our test data set.

B. The SL1MMER algorithm

To solve Eq. (2), an algorithm called SL1MMER, which stands for Scale-down by L1 norm Minimization, Model selection, and Estimation Reconstruction, has been proposed to achieve promising super-resolution power while guaranteeing the efficiency[8][10]. SL1MMER has been originally designed for TomoSAR in urban areas, under the assumption that there are only a few dominant scatterers (phase centers) along elevation axis within each azimuth-range pixel [2]. I.e., $\boldsymbol{\gamma}$ has merely K non-zero entries where typically $K \leq 4$. As its name suggests, this algorithm consists of the following three main steps.

1) Scale-down by L_1 norm minimization: To exploit the sparse prior on $\boldsymbol{\gamma}$, we solve the following L_1-regularized least

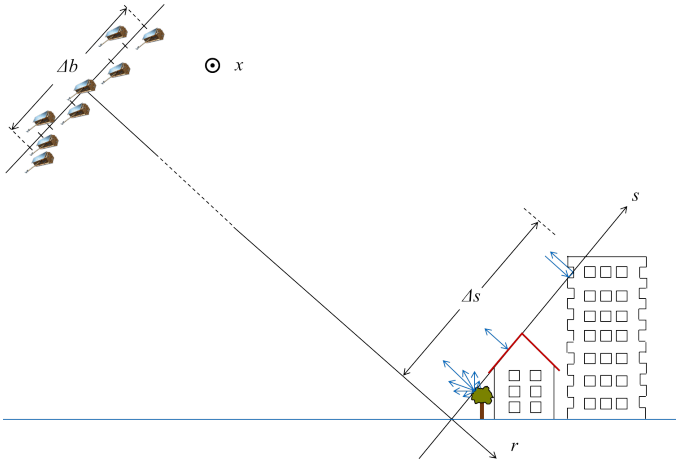


Fig. 6. TomoSAR imaging geometry with an artistic view of TerraSAR-X/TanDEM-X ©DLR. The satellite flies into the plane and looks to its right.

squares problem

$$\hat{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma}} \left\{ \frac{1}{2} \left\| \mathbf{g} - \mathbf{R}\boldsymbol{\gamma} \right\|_2^2 + \lambda_K \left\| \boldsymbol{\gamma} \right\|_1 \right\}, \tag{4}$$

where λ_K is a hyperparameter balancing model error and the sparsity of $\boldsymbol{\gamma}$. Eq. 4 is known to deliver robust elevation estimates $\hat{s}_l$ of dominant scatterers. Therefore, by identifying the most significant entries in $\hat{\boldsymbol{\gamma}}$ and choosing certain columns of $\mathbf{R}$ accordingly, the dimension of the original problem in Eq. (2) can be downscaled by a large factor. However, solving 4 is prone to amplitude bias due to the L_1 norm relaxation. Moreover, outliers might appear when the required mathematical conditions of $\mathbf{R}$ are not fully fulfilled as most of the engineering problems do [32]. These make the next two steps necessary.

2) Model selection: The initial estimate $\hat{\boldsymbol{\gamma}}$ from Eq. (4) may contain artifacts, which falsifies its sparsity level. In order to detect and remove them, the goodness of fit of a model should be penalized by its complexity, so that overfitting of data can be avoided. Model selection can be regarded as the following optimization problem

$$\hat{K} = \arg\min_{K} \left\{ -2 \ln p\left(\mathbf{g} \mid \hat{\boldsymbol{\theta}}(K), K \right) + 2C(K) \right\}, \tag{5}$$

where $p\left(\mathbf{g} \mid \hat{\boldsymbol{\theta}}(K), K \right)$ is the likelihood function of $\mathbf{g}$ given the estimates of unknown $\boldsymbol{\theta}(K)$ and K, $C(K)$ is the penalty term for model complexity. Various alternatives of $C(K)$ have been devised for different needs, e.g., Bayesian information criterion, Akaike information criterion, minimum description length, to name a few [33]. By choosing one specific criterion suitable for the given datasets, Eq. (5) is then solved as combinatorial problem over a pre-defined range of K. Likewise, the most likely positions $\hat{\mathbf{s}}$ of non-zero elements in $\hat{\boldsymbol{\gamma}}$ will hereby be estimated, which further shrinks $\mathbf{R}$. This leaves only one last step to correct amplitude bias.

3) Parameter estimation: At this stage, we have a much slimmer sensing matrix $\mathbf{R}(\hat{\mathbf{s}}) \in \mathbb{C}^{N \times \hat{K}}$. This renders Eq. (2) to

$$\mathbf{g} = \mathbf{R}(\hat{\mathbf{s}})\, \boldsymbol{\gamma}(\hat{\mathbf{s}}) + \mathbf{e}, \tag{6}$$

where $\boldsymbol{\gamma}(\hat{\mathbf{s}}) \in \mathbb{C}^{\hat{K} \times 1}$, and $\mathbf{e} \in \mathbb{C}^{N \times 1}$ is the sum of measurement noise and the error introduced by model selection. Since Eq. (6) is now overdetermined, it can be solved with ordinary least squares (OLS)

$$\hat{\boldsymbol{\gamma}} = \mathbf{R}^{+}(\hat{\mathbf{s}})\, \mathbf{g}, \tag{7}$$

where $(\cdot)^{+}$ denotes pseudo inverse.

Within the framework of SL1MMER, the merits from both sparse regression and OLS have been joined as a whole, namely, robust identification of scatterers' elevation positions, as well as accurate amplitude estimation. Other advantages of SL1MMER over conventional parametric and non-parametric methods have been discussed in [8] and its theoretical limits in terms of estimation accuracy, super-resolution power and the required minimum number of acquisitions for a reasonable reconstruction have been investigated in [10].

C. The M-SL1MMER algorithm

We extend the SL1MMER method to M-SL1MMER, i.e., the multiple-snapshot case. Assume that by applying the method described in Section III, we have already detected M pixels along an iso-height line. We further assume that within each pixel, there is a dominant scatterer located on the considered building façade. Hence, those M scatterers should reside at the same height or elevation position. For each pixel, we have, similar to Eq. (2),

$$\mathbf{g}_m = \mathbf{R}_m \boldsymbol{\gamma}_m + \boldsymbol{\varepsilon}_m, \tag{8}$$

$\forall\, m \in \{1, \ldots, M\}$. If the iso-height line stretches principally in azimuth direction, we expect ξ_n to vary little among all concerned pixels. For this reason, we define $\mathbf{R} := \mathbf{R}_1 \cong \mathbf{R}_2 \cong \ldots \cong \mathbf{R}_M$. By using the identical degree of discretization along elevation axis, we can rewrite Eq. (8) as

$$\mathbf{G} = \mathbf{R}\boldsymbol{\Gamma} + \mathbf{E}, \tag{9}$$

where $\mathbf{G} = [\mathbf{g}_1, \ldots, \mathbf{g}_M]$ is the observation matrix with M measurements vectors, $\boldsymbol{\Gamma} = [\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_M]$ is the unknown discretized reflectivity matrix, and $\mathbf{E}$ accounts for both additive noise and possible model error. Eq. (9) is again an underdetermined system with $N \ll L$ which has infinitely many solutions. Since we assume that all the measurement vectors have one contribution from the same height on a façade, the non-zero entry positions in the columns of $\boldsymbol{\Gamma}$ are aligned in a row-wise fashion. This property of signals is also referred to as joint sparsity. Indeed, there can be more non-zero rows related to ground, lower infrastructures, building roof, etc. Still, the number of non-zeros rows of $\boldsymbol{\Gamma}$ is very limited. To solve Eq. (9) while incorporating this prior, $\hat{\boldsymbol{\Gamma}}$ can be estimated by solving $L_{1,2}$-regularized least squares problem [34],

$$\hat{\boldsymbol{\Gamma}} = \arg\min_{\boldsymbol{\Gamma}} \left\{ \frac{1}{2} \left\| \mathbf{G} - \mathbf{R}\boldsymbol{\Gamma} \right\|_{\mathrm{F}}^2 + \lambda_K \left\| \boldsymbol{\Gamma} \right\|_{1,2} \right\}, \tag{10}$$

where F denotes the Frobenius norm, and the mixed norm $\|\boldsymbol{\Gamma}\|_{1,2} = \sum_{l=1}^{L} \left(\|\boldsymbol{\gamma}^l\|_2 \right)$, where $\boldsymbol{\gamma}^l$ is the l^{th}, promotes joint sparsity. It has been shown in [35] that the probability of successful recovery increases with the number of snapshots.

Note that different polarimetric channels or neighboring pixels were used in a similar way in [36][37].

After the downscaling step based on the estimate in Eq. (10), model selection and parameter estimation will be performed individually for each pixel as the SL1MMER algorithm does.

D. Cramér Rao Lower Bounds (CRLB) for elevation estimates

The Cramér Rao lower bound (CRLB) for elevation estimates $\hat{s}$ for the single-scatterer case has been derived in [38] as

$$\sigma_{\hat{s},0} = \frac{\lambda r}{4\pi\sqrt{N}\cdot\sqrt{2SNR}\cdot\sigma_b}, \tag{11}$$

where σ_b is the standard deviation of b_n. Given an SNR of 3 dB, the CRLB is approx. 1.11 m with all 21 acquisitions.

In urban environment, due to the side-looking geometry of SAR, multiple scatterers are often mapped into one azimuth-range pixel. The fact that closely spaced scatterers will interfere with each other renders a degraded estimation accuracy of individual scatterers [12]. In the interest of super-resolution, the CRLB for elevation estimate of the q^{th} ($q = 1, 2$) scatterer has been derived in [10] as

$$\sigma_{s_q} = c_0 \cdot \sigma_{s_q,0}, \tag{12}$$

where

$$c_0 \approx \max\left\{\sqrt{2.57\left(\alpha^{-1.5} - 0.11\right)^2 + 0.62},\ 1\right\} \tag{13}$$

is the interference factor depending on α which is the distance between two scatterers normalized w.r.t. the Rayleigh resolution unit [10]. c_0 is equal to one (no interference) when two scatterers are far apart, i.e., $\alpha \gg 1$, greater than one since the two scatterer are closely spaced ($\alpha < 1.5$) and increasing with decreasing α.

V. Performance Evaluation Using Simulated Data

In general, as an extension of SL1MMER, M-SL1MMER has the same basic principle. However, instead of exploiting sparsity, M-SL1MMER uses multiple snapshots of iso-height pixels identified in SAR images (with the help of supporting OSM data). Since M-SL1MMER makes use of multiple snapshots and thus more observations than SL1MMER, we naturally expect it to achieve better performance.

In this section, the performance of the proposed M-SL1MMER algorithm, including elevation estimation accuracy, detection rate and false alarm rate in separating overlaid scatterers, and its super-resolution capability, is evaluated using simulated data.

We simulate façade-ground interaction of two scatterers spaced by decreasing elevation distances, which is a well-known TomoSAR benchmark test [2][8]. Note that we only work in the super-resolution regime, i.e., the elevation distance between façade and ground is no larger than the Rayleigh resolution ρ_s. Four scenarios are taken into account with the number of measurements $N \in \{6, 11\}$ and $SNR \in \{3, 10\}$ in [dB] because:

- As mentioned above, eleven is the minimum required number of measurements for a reasonable reconstruction in the interesting SNR range of spaceborne SAR if SL1MMER is used [10];
- In case of two scatterers, six is the number of unknowns, namely the amplitude, phase and elevation positon of each scatterer;
- SNR of 3 dB and 10 dB are usually considered as the lower and upper bound of persistent scatterers, respectively [39].

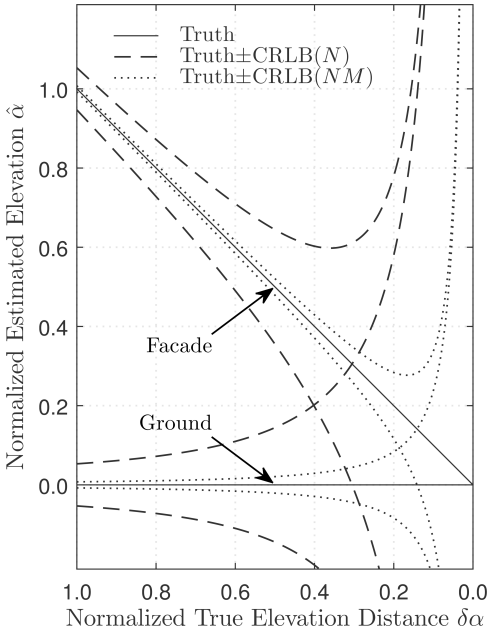


Fig. 7. Normalized true elevation α of simulated façade and ground (solid lines), as well as the CRLB of normalized elevation estimates, both w.r.t. normalized true elevation distance $\delta\alpha$. Dashed lines: CRLB with N baselines; dotted lines: CRLB with NM baselines.

For each façade-ground interaction with a given elevation distance, we independently generate M = 48 snapshots, which is an average case for the test buildings in Figure 5. The true elevation of simulated façade and ground is shown as two solid line segments w.r.t. their normalized elevation distance $\delta\alpha$ in Figure 7, respectively. In addition, we show for the case N = 11 the evolution of CRLB, which increases with decreasing elevation distance of two interfering scatterers in the super-resolution regime, as implied by 12 and 13. The dashed lines mark true elevation $\pm 1 \times \text{CRLB}$ with N interferograms, denoted as $\text{CRLB}(N)$, while we plot true elevation $\pm 1 \times \text{CRLB}$ with NM interferograms, denoted as $\text{CRLB}(NM)$, as dotted lines. We will show that M-SL1MMER using N interferograms and M snapshots approaches the latter bound, which can be achieved by SL1MMER given NM interferograms.

We solve the L_1- and $L_{1,2}$-regularized least squares problems independently, and then follow the SL1MMER procedures to perform model selection and parameter estimation. λ_K is chosen adaptively, which depends on N, M and the noise level of observations [34]. The results are shown in Figure 8 and Figure 9 with elevation estimates $\hat{\alpha}$ of façade and ground plotted w.r.t. their normalized true elevation difference $\delta\alpha$. Each dot depicts mean value of all estimates, with error bar indicating its standard deviation. In each subplot, the two solid line segments mark the true elevation for façade

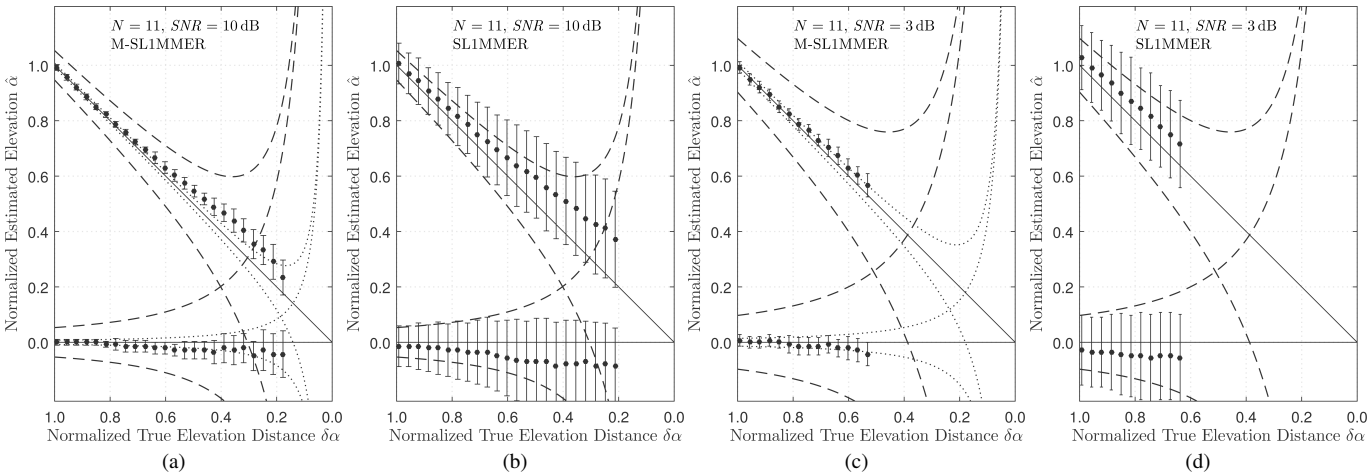


Fig. 8. Reconstructed elevation of simulated façade and ground with M = 48, N = 11; (a) SNR = 10 dB with M-SL1MMER (b) SNR = 10 dB with SL1MMER, (c) SNR = 3 dB with M-SL1MMER and (d) SNR = 3 dB with SL1MMER respectively. Each dot has the sample mean of all estimates as its y value and the corresponding standard deviation as error bar.

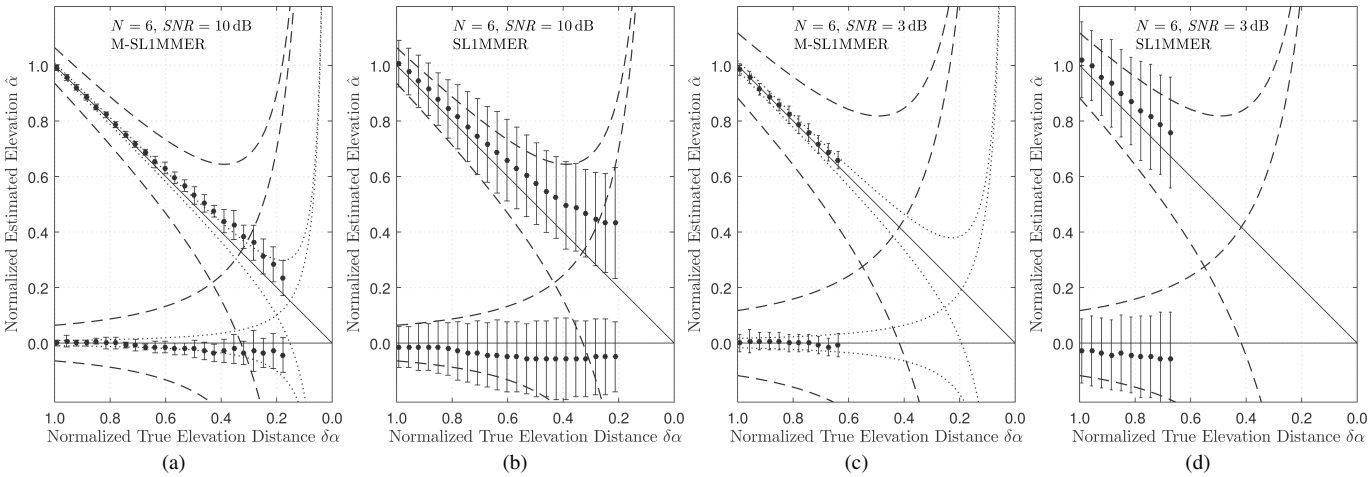
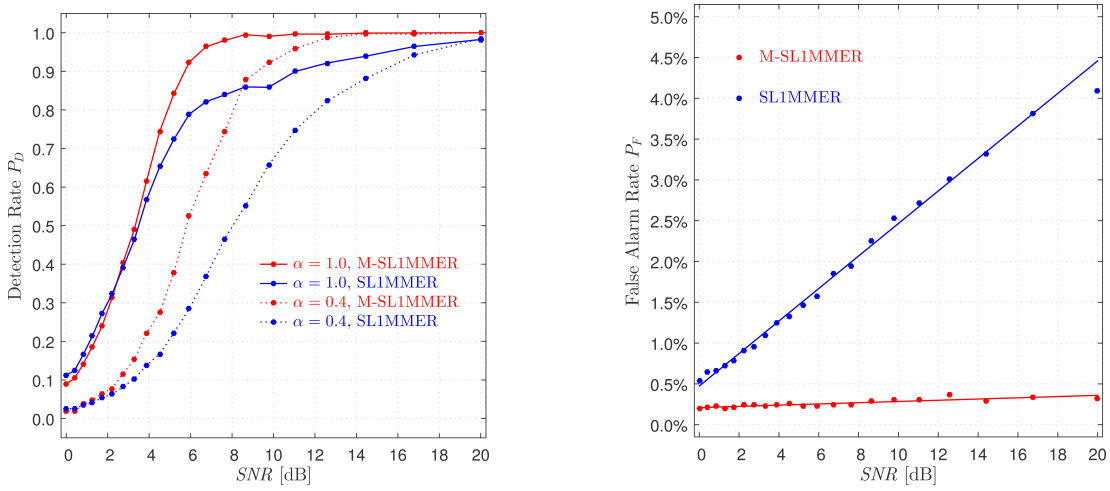


Fig. 9. Reconstructed elevation of simulated façade and ground with M = 48, N = 6, (a) SNR = 10 dB with M-SL1MMER, (b) SNR = 10 dB with SL1MMER, (c) SNR = 3 dB with M-SL1MMER, and (d) SNR = 3 dB with SL1MMER, respectively. Each dot has the sample mean of all estimates as its y value and the corresponding standard deviation as error bar.

and ground, respectively, while the dashed and dotted lines denote true elevation $\pm 1 \times$ CRLB, which is the same as in Figure 7. Missing points suggest that detection rate is below 25%. Note that we call detection when not only two scatterers are separated, but also their estimates should be bounded by $\pm 3 \times \mathrm{CRLB}(N)$ of their true elevation. For N = 11, the elevation estimates with both methods are still somewhat comparable, despite the fact that joint sparsity model leads to much smaller variance and slightly better super-resolution capability. SL1MMER performs in particular worse with smaller N and lower SNR. On the contrary, even for the case N = 6, reasonable elevation has been reconstructed with M-SL1MMER. In particular, M-SL1MMER with N interferograms and M snapshots is equivalent to SL1MMER with NM measurements in relation to the standard deviation of elevation estimates, although the mean values slowly drift away from the true elevation with increasing interference between two scatterers.

In Figure 10, the detection rate P_D is provided for the case N = 11 w.r.t. normalized true elevation distance $\delta\alpha$. The red and blue colors denote M-SL1MMER and SL1MMER, respectively. The solid and dashed lines illustrate the results with SNR = 10 dB and 3 dB, respectively. If we define elevation resolution to be the minimum distance between façade and ground required to achieve at least 50% detection rate, then the resolution of M-SL1MMER is approximately one tenth of Rayleigh resolution better than the one of SL1MMER, given an SNR of 10dB before. We also analyzed P_D w.r.t. SNR for two typical elevation distances α = 1 or 0.4, i.e., when two scatterers are spaced by one or two fifths of Rayleigh resolution. The results are shown in Figure 11(a). Moreover, false alarm rate P_F is illustrated in Figure 11(b) as a function of SNR for M-SL1MMER (red) and SL1MMER (blue), respectively.In this context, we simulate only one scatter and define false alarm for the case when two scatterers are detected. These analyses confirm the fact that M-SL1MMER outperforms SL1MMER in all respects significantly. For the case N = 6, the gain of using multiple snapshots regarding



(a) (b)

Fig. 11. Detection rate P_D and false alarm rate P_F w.r.t. SNR, for the case $N = 11$: (a) P_D for $\alpha = 1.0$ with M-SL1MMER (red), $\alpha = 1.0$ with SL1MMER (orange), $\alpha = 0.4$ with M-SL1MMER (violet), and $\alpha = 0.4$ with SL1MMER (blue), respectively. (b) P_F with M-SL1MMER (red), and SL1MMER (orange), respectively.

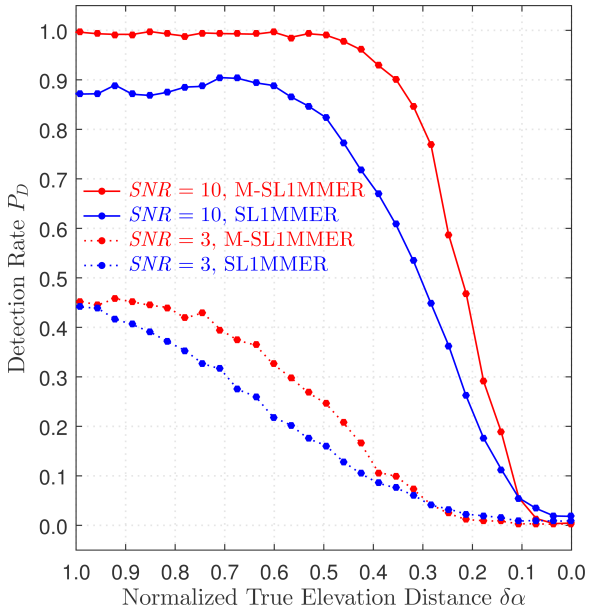


Fig. 10. Detection rate P_D w.r.t. normalized true elevation distance $\delta\alpha$ between façade and ground, for the case $N = 11$. Red: $SNR = 10$ dB with M-SL1MMER, orange: $SNR = 10$ dB with SL1MMER, violet: $SNR = 3$ dB with M-SL1MMER, and $SNR = 3$ dB with SL1MMER, respectively.

P_D and P_F is comparable to the case of $N = 11$.

VI. Practical Demonstration Using TanDEM-X Data

In this section, M-SL1MMER is applied to the TanDEM-X data mentioned in Section II. The results are compared to those obtained using SL1MMER. Figure 12–14 show the reconstructed and color-coded elevation of the two test buildings in Figure 5, overlaid with intensity. From Figure 12 to 14, 21, 11 and 6 interferograms are used, respectively. The separated superimposed scatterers, from left to right estimated using M-SL1MMER and SL1MMER and from top to down of first and second layer, are illustrated respectively.

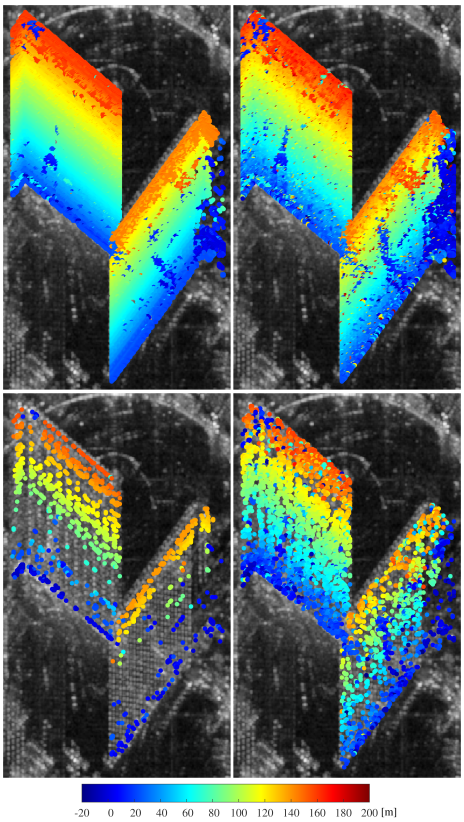


Fig. 12. Reconstructed and color-coded elevation of the two test buildings using 21 interferograms, visualized in two layers, overlaid with intensity. From top to down: first and second layer, respectively; from left to right: M-SL1MMER and SL1MMER, respectively.

On top of the test buildings, reflections from building roof and façade are overlaid. In these practical examples, dominating scattering from roof (dark red) can be seen in the first layer, whereas the corresponding parts of façade (light red) are

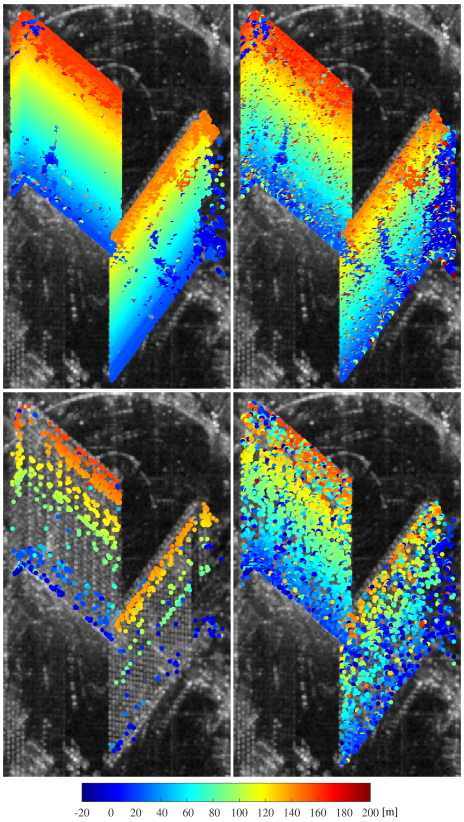


Fig. 13. The same results as Figure 12, but using 11 interferograms.

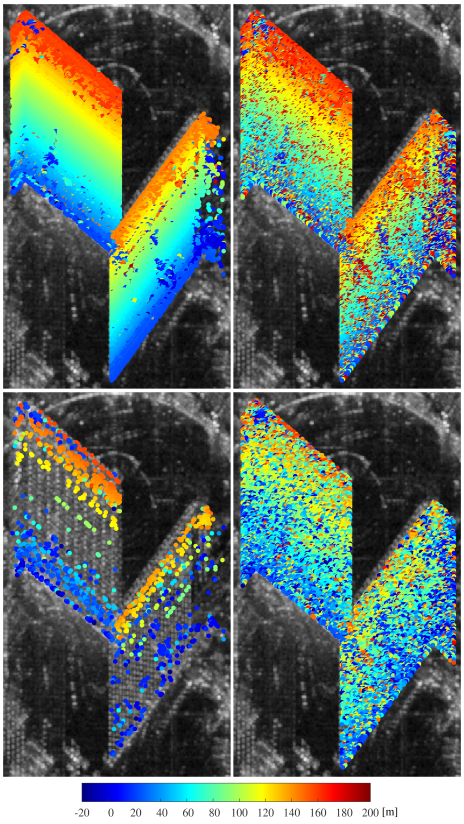


Fig. 14. The same results as Figure 12, but using 6 interferograms only.

visible in the second layer. We do not expect many reflections from lower structures though, due to the large slope of the shell-like roof in front of the test buildings. It is evident that M-SL1MMER (left) using joint sparsity model significantly outperforms SL1MMER (right). In particular, when $N = 6$, i.e., using extremely small number of scenes, the second layer estimated using SL1MMER (lower right plot of Figure 14) is deteriorated by false alarms while M-SL1MMER still achieves reasonable results (lower left plot of Figure 14).

In particular, due to the significantly improved estimation accuracy, M-SL1MMER reconstructs some interesting details which was not accessible so far. For a practical demonstration, we calculated elevation distance between first and second layer for the double-scatterer case, which is shown in Figure 15. The red parallelogram marks the area where facade and roof are overlaid, cf. Figure 16. At the far-range side of this area, the elevation distance amounts to approximately 22.60 [m] (cyan). Accordingly, the width of the roof can be calculated to be 18.27 [m], which agrees, up to the decimeter level, with what we estimated from the 3-D building model of Google Earth. Besides, the yellow parallelogram circumfuses the area where two neighboring windows in the diagonal direction, exemplified as S_1 and S_2 in Figure 16, are superimposed. Thus their elevation distance is more or less constant. This can be easily verified by comparing the SAR amplitude image to the optical one in Figure 16. As can be counted from Figure 16, the number of windows per floor adds up to 20, whereas there are only 10 extraordinarily bright pixels in the SAR amplitude image. In this area, the results with M-SL1MMER are much more homogeneous, which indicates a more robust reconstruction.

VII. Concluding Remarks

In this paper, a novel framework is proposed which can achieve precise TomoSAR reconstruction while significantly reducing the required number of images. The core idea is the exploitation of joint sparsity in iso-height SAR pixel groups that can be identified with the support of online available GIS data—2D building footprints. Experiments using bistatic TanDEM-X data stacks demonstrate the great potential of the proposed approach.

A few additional remarks might be helpful for further use of our results:

- The approach we proposed for grouping iso-height pixels can also be used for all other estimators that support multiple-snapshot (also referred to as multi-look) estimation, like NLS, MUSIC, etc.;
- Our experiments are performed over Las Vegas where moderate quality GIS data are available. For test areas with relatively poor 2-D footprint data, the benefit of the proposed approach that refine the iso-height models of each building using, e.g., SAR intensity map, could be even more evident;
- In the cities where the GIS data are not available, one can use the 2-D footprint reconstructed using a preliminarily retrieved TomoSAR point cloud [40] to alternately improve the TomoSAR estimation procedure;

- The proposed M-SL1MMER is a general spectral estimator, even though we applied it here for tomographic reconstruction;
- In our work, the mix $L_{1,2}$ norm is introduced to promote the joint sparsity prior. Depending on the applications, this constraint can be relaxed by replacing $L_{1,2}$ with $L_{p,2}$ norm with $p > 1$;

In the future, we will extend the proposed M-SL1MMER for higher dimensional spectral estimation problems, e.g., differential tomographic SAR reconstruction using mixed single- and multi-pass monostatic data stacks.

ACKNOWLEDGMENT

This work is supported by the Helmholtz Association under the framework of the Young Investigators Group "SiPEO" (VH-NG-1018, www.sipeo.bgu.tum.de) and International Graduate School of Science and Engineering, Technische Universität München (Project 6.08: "4D City").

REFERENCES

[1] S. Gernhardt, N. Adam, M. Eineder, and R. Bamler, "Potential of Very High Resolution SAR for Persistent Scatterer Interferometry in Urban Areas," *Annals of GIS*, vol. 16, no. 2010-06, pp. 103–111, 2010.

[2] X. X. Zhu and R. Bamler, "Very High Resolution Spaceborne SAR Tomography in Urban Environment," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 12, pp. 4296–4308, Dec. 2010.

[3] D. Reale, G. Fornaro, A. Pauciullo, X. Zhu, and R. Bamler, "Tomographic Imaging and Monitoring of Buildings With Very High Resolution SAR Data," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 4, pp. 661–665, Jul. 2011.

[4] G. Fornaro, A. Pauciullo, D. Reale, X. Zhu, and R. Bamler, "Sar tomography: an advanced tool for 4d spaceborne radar scanning with application to imaging and monitoring of cities and single buildings," *IEEE Geosci. Remote Sensing Newslett*, pp. 10–18, 2012.

[5] Y. Wang, X. X. Zhu, Y. Shi, and R. Bamler, "Operational tomosar processing using terrasar-x high resolution spotlight stacks from multiple view angles," in *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*. IEEE, 2012, pp. 7047–7050.

[6] X. X. Zhu and M. Shahzad, "Facade Reconstruction Using Multiview Spaceborne TomoSAR Point Clouds," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 6, pp. 3541–3552, Jun. 2014.

[7] G. Fornaro, F. Lombardini, A. Pauciullo, D. Reale, and F. Viviani, "Tomographic Processing of Interferometric SAR Data: Developments, applications, and future research perspectives," *IEEE Signal Processing Magazine*, vol. 31, no. 4, pp. 41–50, Jul. 2014.

[8] X. X. Zhu and R. Bamler, "Tomographic SAR Inversion by L1-Norm Regularization: The Compressive Sensing Approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 10, pp. 3839–3846, Oct. 2010.

[9] A. Budillon, A. Evangelista, and G. Schirinzi, "Three-Dimensional SAR Focusing From Multipass Signals Using Compressive Sampling," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 1, pp. 488–499, Jan. 2011.

[10] X. X. Zhu and R. Bamler, "Super-Resolution Power and Robustness of Compressive Sensing for Spectral Estimation With Application to Spaceborne Tomographic SAR," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 1, pp. 247–258, Jan. 2012.

[11] ——, "Demonstration of Super-Resolution for Tomographic SAR Imaging in Urban Environment," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 8, pp. 3150–3157, Aug. 2012.

[12] ——, "Superresolving SAR Tomography for Multidimensional Imaging of Urban Areas: Compressive sensing-based TomoSAR inversion," *IEEE Signal Processing Magazine*, vol. 31, no. 4, pp. 51–58, Jul. 2014.

[13] A. Ferretti, C. Prati, and F. Rocca, "Analysis of Permanent Scatterers in SAR interferometry," in *Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 International*, vol. 2, 2000, pp. 761–763 vol.2.

[14] B. M. Kampes, *Radar Interferometry: Persistent Scatterer Technique*. Springer Science & Business Media, Sep. 2006.

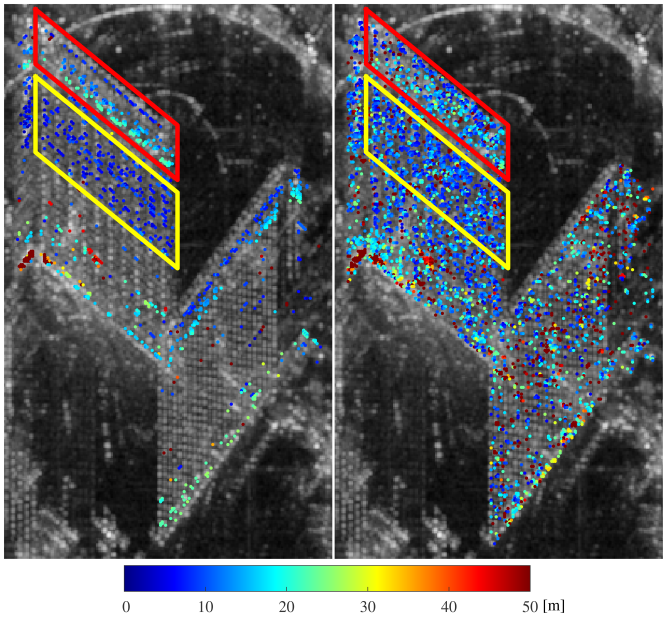


Fig. 15. Elevation distance between first and second layer with $N = 21$. Left: M-SL1MMER; right: SL1MMER.

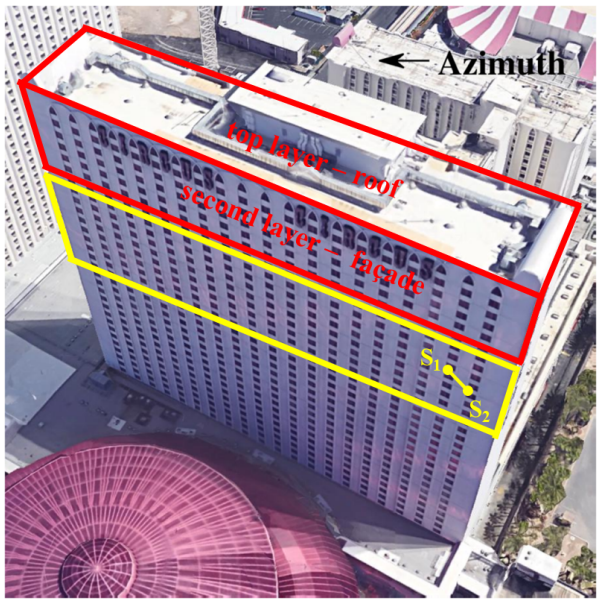


Fig. 16. Optical image of the left test building in Figure 15 ©Google.

[15] N. Adam, B. Kampes, M. Eineder, J. Worawattanamateekul, and M. Kircher, "The development of a scientific permanent scatterer system," in *ISPRS Workshop High Resolution Mapping from Space, Hannover, Germany*, 2003.

[16] G. Fornaro, F. Lombardini, and F. Serafino, "Three-dimensional multipass SAR focusing: experiments with long-term spaceborne data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 702–714, Apr. 2005.

[17] F. Lombardini, "Differential tomography: a new framework for SAR interferometry," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 1, pp. 37–44, Jan. 2005.

[18] A. Ferretti, A. Fumagalli, F. Novali, C. Prati, F. Rocca, and A. Rucci, "A new algorithm for processing interferometric data-stacks: Squeesar," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 9, pp. 3460–3470, Sept 2011.

[19] Y. Wang, X. X. Zhu, and R. Bamler, "Retrieval of phase history parameters from distributed scatterers in urban areas using very high resolution sar data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 73, pp. 89 – 99, 2012.

[20] M. Schmitt and U. Stilla, "Adaptive multilooking of airborne single-pass multi-baseline insar stacks," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 1, pp. 305–312, Jan 2014.

[21] R. Guo and X. X. Zhu, "High-Rise Building Feature Extraction Using High Resolution Spotlight TanDEM-X Data," in *EUSAR 2014; 10th European Conference on Synthetic Aperture Radar; Proceedings of*, Jun. 2014, pp. 1–4.

[22] "Stats - OpenStreetMap Wiki." [Online]. Available: http://wiki.openstreetmap.org/wiki/Statistics

[23] H. Fan, A. Zipf, Q. Fu, and P. Neis, "Quality Assessment for Building Footprints Data on OpenStreetMap," *Int. J. Geogr. Inf. Sci.*, vol. 28, no. 4, pp. 700–719, Apr. 2014.

[24] M. Haklay, "How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets," *Environment and Planning B: Planning and Design*, vol. 37, no. 4, pp. 682–703, 2010.

[25] M. Helbich, C. Amelunxen, and P. Neis, "Comparative Spatial Analysis of Positional Accuracy of OpenStreetMap and Proprietary Geodata," in *AGILE 2011*, 2011, gIScience, Department of Geography, University of Heidelberg, Berliner Strasse 48, 69120 Heidelberg, Germany.

[26] P. Mooney, P. Corcoran, and A. C. Winstanley, "Towards Quality Metrics for OpenStreetMap," in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ser. GIS '10. New York, NY, USA: ACM, 2010, pp. 514–517.

[27] C. K. Robert Hecht, "Measuring Completeness of Building Footprints in OpenStreetMap over Space and Time," *ISPRS International Journal of Geo-Information*, vol. 2, pp. 1066–1091, 2013.

[28] A. Reigber and A. Moreira, "First demonstration of airborne sar tomography using multibaseline l-band data," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 38, no. 5, pp. 2142–2152, Sep 2000.

[29] G. Fornaro, F. Serafino, and F. Soldovieri, "Three-dimensional focusing with multipass SAR data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 3, pp. 507–517, Mar. 2003.

[30] G. Fornaro, D. Reale, and F. Serafino, "Four-Dimensional SAR Imaging for Height Estimation and Monitoring of Single and Double Scatterers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 1, pp. 224–237, Jan. 2009.

[31] X. X. Zhu and R. Bamler, "Let's Do the Time Warp: Multicomponent Nonlinear Motion Estimation in Differential SAR Tomography," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 4, pp. 735–739, Jul. 2011.

[32] X. X. Zhu, "Very high resolution tomographic sar inversion for urban infrastructure monitoringa sparse and nonlinear tour," Ph.D. dissertation, Technische Universität München, 2011.

[33] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed. New York: Springer, Dec. 2003.

[34] D. Malioutov, M. Cetin, and A. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[35] Y. Eldar and H. Rauhut, "Average Case Analysis of Multichannel Sparse Recovery Using Convex Relaxation," *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 505–519, Jan. 2010.

[36] E. Aguilera, M. Nannini, and A. Reigber, "Multi-signal compressed sensing for polarimetric SAR tomography," in *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, Jul. 2011, pp. 1369–1372.

[37] M. Schmitt and U. Stilla, "Compressive Sensing Based Layover Separation in Airborne Single-Pass Multi-Baseline InSAR Data," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 2, pp. 313–317, Mar. 2013.

[38] R. Bamler, M. Eineder, N. Adam, X. Zhu, and S. Gernhardt, "Interferometric Potential of High Resolution Spaceborne SAR," *Photogrammetrie - Fernerkundung - Geoinformation*, vol. 2009, no. 5, pp. 407–419, Nov. 2009.

[39] N. Adam, R. Bamler, M. Eineder, and B. Kampes, "Parametric estimation and model selection based on amplitude-only data in ps-interferometry," in *ESA FRINGE Workshop*, 2005.

[40] M. Shahzad and X. X. Zhu, "Robust reconstruction of building facades for large areas using spaceborne tomosar point clouds," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 53, no. 2, pp. 752–769, Feb 2015.